Image and Vision Computing 37 (2015) 1–11

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Robust stereo matching using adaptive random walk with restart algorithm☆

Sehyung Lee [a], Jin Han Lee [a], Jongwoo Lim [b,*], Il Hong Suh [a]

[a] Department of Electronics and Computer Engineering, Hanyang University, 133-791 Seoul, South Korea
[b] Division of Computer Science and Engineering, Hanyang University, 133-791 Seoul, South Korea

ARTICLE INFO

Article history:
Received 9 November 2013
Received in revised form 4 October 2014
Accepted 9 January 2015
Available online 20 February 2015

Keywords:
Global optimization
Random walk with restart
Stereo matching
Superpixels

ABSTRACT

In this paper, we propose a robust dense stereo reconstruction algorithm using a random walk with restart. The pixel-wise matching costs are aggregated into superpixels and the modified random walk with restart algorithm updates the matching cost for all possible disparities between the superpixels. In comparison to the majority of existing stereo methods using the graph cut, belief propagation, or semi-global matching, our proposed method computes the final reconstruction through the determination of the best disparity at each pixel in the matching cost update. In addition, our method also considers occlusion and depth discontinuities through the visibility and fidelity terms. These terms assist in the cost update procedure in the calculation of the standard smoothness constraint. The method results in minimal computational costs while achieving high accuracy in the reconstruction. We test our method on standard benchmark datasets and challenging real-world sequences. We also show that the processing time increases linearly in relation to an increase in the disparity search range.

1. Introduction

Depth perception is a fundamental issue in the study of computer vision. Depth information is used in object recognition, human tracking, and image segmentation, as well as in robotics applications, including navigation, localization, and mapping. The depth is therefore an important measure in order to understand space and the objects within it. The stereo camera is a popular method to measure a 3D point cloud. Stereo matching algorithms attempt to determine corresponding objects between the two scenes. There has been considerable progress in this field. Scharstein and Szeliski [1] showed that the majority of stereo matching algorithms can be built from four basic components:

- Computation of local matching costs for all pixels.
- Aggregation of pixel-wise matching costs in the support regions.
- Search for the global optimal disparity values.
- Refinement of the resultant disparity map.

The stereo algorithms can be classified into two groups, local and global, depending on whether the global search and refinement are performed. Initially, the pixel-wise matching costs between the reference and target images are calculated. The pixel-wise costs are often noisy, and contain minimal information in texture-less regions. Therefore, the costs of neighboring support regions are aggregated together. A local algorithm would finish at this stage and assign the best disparity value based on the aggregated costs at each pixel. Thus, local methods are typically faster, yet less accurate in comparison to a global technique.

In contrast, a global method takes account of the overall scene structure and depth smoothness in order to determine the disparity map. Recently, stereo algorithms have achieved impressive results using the Markov Random Field (MRF). The labeling of an MRF is known to be NP-hard. Due to this factor, belief-propagation (BP) and graph cut (GC) methods are widely used to approximate the optimal solution of the global method. However, BP- or GC-based stereo matching algorithms require substantial computational resources. In this study, we present a novel stereo matching algorithm that utilizes a random walk with restart (RWR) to optimize the matching cost without the need for the conventional BP or GC algorithms. The proposed method provides two benefits: computational efficiency, and theoretical optimality. These factors enable the proposed method to achieve high-quality matching results at relatively low computational cost.

We also propose an adaptive RWR algorithm (ARW) that is suited for practical applications. The use of a stereo camera presents many problems, including low texture, occlusion, and depth discontinuities. A general energy minimization formulation contains data and smoothness terms, which are effective when considering the low-texture areas of an image. However, this approach will often fail when conserving occlusion or depth discontinuities, resulting in an overall unsatisfactory matching. In order to solve this problem, our proposed algorithm includes two additional procedures. Firstly, the occluded regions are detected through

☆ This paper has been recommended for acceptance by Enrique Dunn.
* Corresponding author. Tel.: +82 2 2220 2376.
E-mail addresses: shl@incorl.hanyang.ac.kr (S. Lee), jhlee@incorl.hanyang.ac.kr (J.H. Lee), jlim@hanyang.ac.kr (J. Lim), ihsuh@hanyang.ac.kr (I.H. Suh).

http://dx.doi.org/10.1016/j.imavis.2015.01.003
0262-8856/

a left-right consistency check. Secondly, we introduce a robust penalty function to compute an additional fidelity term that allows us to preserve the depth discontinuity. These two advances play an important role in improving the performance.

This paper's primary contribution lies in its introduction of a novel global optimization method for stereo matching. Our proposed algorithm is based on the RWR method, instead of the conventional GC or BP. We describe a new interpretation of the RWR method, and outline its benefits, in relation to stereo matching and classical optimization algorithm. In addition, we modify the conventional RWR algorithm for our purposes, by taking account of regions of occlusion and discontinuity. With these modifications, we achieve high-accuracy stereo reconstruction at little computational cost in comparison to GC- and BP-based methods. This is our main contribution to the literature. Our experimental results quantitatively demonstrate that the performance of our method is comparable to existing state-of-the-art techniques. We also implement our proposed method in a real-world test environment, and reveal its minimal processing time requirements.

2. Related work

We mentioned in the preceding section that stereo matching algorithms can be classified as local or global. Some local methods are comparable to global methods with respect to their accuracy. Yoon and Kweon [2] proposed an adaptive weighting method similar to a bilateral filter because of the way the filtered result is a weighted average of neighboring pixels. Their algorithm preserves depth edge information, yet the full-kernel implementation is very slow. Rhemann [3] proposed a cost–volume filtering algorithm, an approximation of the bilateral filter, and Yang [4] improved the cost–volume filtering to be more efficient. These approaches both require a substantial window size in order to accurately estimate the disparity. This increases their computational cost in comparison to other local methods such as that developed by Geiger [5]. De-maeztu [6] proposed linear stereo matching in order to overcome these weaknesses.

In comparison to local methods, a global algorithm will typically skip the cost aggregation step, and instead aim to minimize a global cost function consisting of a data and a smoothness term. Significant advances have been made in global optimization-based stereo matching algorithms over the past decade. The majority of global stereo matching algorithms treat the problem as one of energy minimization, which is often formulated through inference on a MRF. The problem of finding an exact optimal solution in a looping MRF is known to be NP-hard. To alleviate this factor, approximation methods, such as GC- [7] or BP-based stereo matching [8] have been proposed. These algorithms have been extended further, including the integration of multiple cues, such as segmentation [9–11], and left-right consistency [12]. The multiple cues are typically used to address occlusion, depth discontinuity, and low-texture areas. Yang [13] proposed a color-weighted hierarchical BP algorithm. This allows global matching algorithms to account for depth discontinuities using segmentation as a soft constraint, and was shown to converge quickly.

In comparison, stereo matching algorithms, including both local and global methods, usually assume a fronto-parallel plane in the window-based matching cost aggregation stage. There have been several algorithms proposed to deal with the limitation of the fronto-parallel assumption. Li and Zucker [14] introduced surface geometric constraints in the global optimization framework. In contrast, Bleyer [15] proposed a PatchMatch-based stereo matching algorithm, while Yamaguchi [16] developed a slanted-plane MRF-based algorithm.

However, almost all existing high-performance algorithms are limited in real-time implementation. A semi-global matching (SGM) algorithm has been proposed in order to reduce the optimization complexity of the global cost function. Hirschmuller [17] presented a SGM algorithm that approximated a global cost function using path-wise optimizations. In recent years, modified algorithms based on SGM have been proposed [18]. SGM methods have been successfully applied to stereo matching.

MRF-based energy minimization has been used in stereo matching algorithms for the last decade. The GC, BP, and SGM algorithms generally form the basis of current state-of-the-art methods. In this paper, we instead employ the RWR algorithm to optimize the matching cost. Several applications have successfully used RWR, including Google's famous PageRank algorithm [19], content-based image retrieval [20], and image segmentation [21]. However, RWR is not as widely used in the stereo matching problem as GC- and BP-based methods.

3. Algorithm description

In this section, we describe our proposed stereo matching algorithm. In Fig. 1, an overview of the proposed algorithm is illustrated. The method computes the initial matching cost for each pixel and all disparities using a combination of a census transform and gradient image matching. We describe this further in Section 3.1. The cost of matching is aggregated into the superpixel that each pixel belongs to. We describe this process further in Section 3.2. Finally, the aggregated matching cost is updated via the RWR algorithm to determine the optimum disparity. Our modified RWR algorithm takes into account the occluded and discontinuity regions based on our current understanding. This is described further in Section 3.3. Our proposed method uses gray-scale images, as well as Gaussian smoothing ($3 \times 3, \sigma = 1.0$) to reduce minor noise.

3.1. Robust local matching

In this stage, the initial pixel-wise matching costs are calculated. Local matching methods are used to develop a measure that provides optimal high-quality matching. Of these, the gradient image, census transform, rank transform, and mutual information are known to be both robust and accurate. In our local matching algorithm, we use both census transform and gradient image matching as inspiration.

The census-based matching technique was originally introduced by Zabi and Woodfill [22]. The pixels of both left and right images are transformed into a binary vector and compared to the surrounding pixels within finite support regions:

$$T(u,v) = \bigotimes_{(u_w,v_w)\in \mathrm{w}(u,v)} H(I(u,v), I(u_w,v_w)), \tag{1}$$

where $I(u, v)$ and $I(u_w, v_w)$ denote the intensity values of the input pixel and the surrounding pixels, respectively; $\otimes$ denotes concatenation, w is the window around (u, v), and H is a binary function returning 0 or 1. We use a 5×5 window in order to encode each pixels binary vector in the census transform. The binary vector is encoded by comparing the intensity values of a center and its surrounding pixels:

$$H(I(u,v), I(u_w,v_w)) = \begin{cases} 0 \text{ if } & I(u,v) < I(u_w,v_w), \\ 1 \text{ if } & I(u,v) \geq I(u_w,v_w). \end{cases} \tag{2}$$

The binary vector is assigned to each pixel in the left and right images. The matching cost is computed using the Hamming distance of two binary vectors:

$$\begin{aligned} C_r(u,v,d) &= Hamming(T_l(u,v), T_r(u+d,v)), \\ C_l(u,v,d) &= Hamming(T_r(u,v), T_l(u-d,v)), \end{aligned} \tag{3}$$

where $C(u, v, d)$ is the Hamming distance based cost function at disparity d, and the subscripts l and r denote the left and right images, respectively. The census transformation is robust to radiometric variations and image noise owing to the local image structures being encoded based on the relative ordering of pixel intensities. However, matching ambiguities can be caused in some regions as a result of this property; for example, repetitive or similar texture patterns are a particular problem. In order